100. Os processos de auditoria baseiam-se muitas vezes em procedimentos estatísticos. Algumas auditorias exaustivas aos gastos semanais em alimentação numa enfermaria forneceram os seguintes valores (em centenas de contos): 120, 125, 128, 130, 110, 128, 129, 131, 125, 127. Se supusermos que o valor gasto é descrito por uma variável aleatória $\mathcal{N}(\mu, 10)$, qual a probabilidade de se observar um gasto de pelo menos 135? E de menos do que 115? O que dizer destes valores se a distribuição for $\mathcal{N}(125, \sigma^2)$?

101. O tempo necessário à manufacturação de uma peça foi medido quando esta era executada por 50 trabalhadores, obtendo-se um tempo médio de execução de 3.2 horas e uma variância amostral de 8.9. Acha que será lícito supor que a variável aleatória que descreve o tempo de manufacturação é exponencial? E, neste caso, qual o respectivo parâmetro? E se a variância da amostra fosse igual a 29.8, acha razoável manter a aproximação por uma distribuição exponencial?

102. O número de passageiros registado em 35 autocarros num período de 1 hora foi:

200	221	200	290	220	250	227
225	240	210	230	222	250	238
223	250	300	328	224	245	240
230	223	220	230	230	240	233
228	250	225	229	250	245	235

Parece-lhe razoável ajustar uma distribuição normal ou uma distribuição exponencial a estes dados? Que lhe parece o ajuste de uma distribuição de Poisson? Em cada um dos casos que julgar pertinentes, calcule os respectivos parâmetros.

103. Hoje em dia é corrente os obstetras avançarem com um prognóstico para a data de nascimento logo no início da gravidez. Naturalmente, nem sempre acertam, verificando-se desvios quer por antecipação quer por atraso no nascimento. Uma observação destes desvios em 1000 partos forneceu uma média de 5.3 dias de atraso com uma variância amostral de 8.4. Pretende-se descrever a variável aleatória que representa, em dias, o desvio de um parto relativamente à data do prognóstico do obstetra. Parece-lhe razoável supor que esta variável tem uma distribuição exponencial? E uma distribuição normal? Calcule a probabilidade do parto ocorrer pelo menos 10 dias após a data de prognóstico. Calcule a probabilidade de o parto ocorrer pelo menos 5 dias antes do prognosticado. Qual a probabilidade de o prognóstico falhar por mais do que 3 dias?

104. Uma análise clínica efectuada a 50 pacientes utilizava duas drogas, das quais se pretende saber qual o efeito sobre um parâmetro em estudo (denominado o HAM-D). Os pacientes foram divididos em 2 grupos, sendo aplicada a cada um destes apenas um dos produtos em estudo. Obtiveram-se os seguintes registos no nível de HAM-D.

Grupo 1												
6	13	21	7	4	12	16	9	11	19	7	20	24
10	20	21	8	24	24	14	12	2	15	16	16	
Grupo 2												
23	8	9	8	12	6	6	5	17	12	14	14	16
14	17	3	11	13	8	14	5	16	15	16	5	

Que tipo de distribuição lhe parece mais correcto para descrever estes dados? Quais os valores dos parâmetros da família de distribuições escolhida? Os dois grupos parecem-lhe diferentes? Para cada grupo, que probabilidades calcula para a observação de um nível inferior a 10 do HAM-D?

105. Uma variável aleatória tem densidade $f(x) = \frac{x-\alpha}{2} \mathbb{I}_{[\alpha,\alpha+2]}(x)$ onde $\alpha \in \mathbb{R}$ é desconhecido. Se dispusermos de uma amostra $X_1, \ldots, X_n$ dessa variável aleatória, qual o estimador do máximo de verosimilhança para α? E o estimador do método dos momentos?

106. Se X tiver densidade $f(x) = \lambda e^{-\lambda x} \mathbb{I}_{[0,+\infty)}(x)$ e utilizarmos para estimar o parâmetro a estatística $d_1(X_1, \ldots, X_n) = \max(X_1, \ldots, X_n)$. Este estimador é cêntrico? Qual o valor do risco que lhe está associado?

107. A densidade indicada no exercício anterior implica que $\mathrm{E}(X) = \mathrm{Var}(X) = \frac{1}{\lambda}$. Se os valores seguintes constituirem uma amostra para X, indique duas aproximações para o parâmetro λ.

1.147	0.333	1.597	4.283	1.090	0.376	3.023	0.646	7.776	0.147
1.414	3.525	7.207	1.775	1.360	0.795	5.567	3.667	0.016	1.970

108. Um problema que torna não trivial o tratamento de algumas questões ditas *delicadas* (como, por exemplo, questões sobre uso de drogas, hábitos sexuais, doenças socialmente mal aceites, etc.) deve-se ao facto de algumas respostas não serem verdadeiras. Para fixarmos ideias, vamos supor que perguntamos a cada pessoa se é do tipo **Q** (representando **Q** o atributo ou doença que pretendemos estudar). Cada pessoa deverá responder com *sim* ou *não*. A questão directa, *é uma pessoa do tipo* **Q**?, arrisca-se a obter respostas falsas, conforme já indicado, ou recusas à formulação de resposta. Para tentar tornear estas dificuldades propõe-se um esquema em que a pergunta directa poderá ou não aparecer. Um esquema possível é o seguinte: primeiro a pessoa inquirida lança um dado; se sair face 5 ou 6, deve responder à pergunta *é uma pessoa do tipo* **Q**?; se, no lançamento do dado sair 1, 2, 3 ou 4, deverá responder a uma qualquer pergunta que não levante problemas, por exemplo, *sabe quem foi Monet?* (pintor impressionista que viveu entre 1840 e 1926). Ambas as perguntas podem ser respondidas com *sim* ou *não* e, com o lançamento do dado feito pelo inquirido, quem recebe a resposta não sabe qual das perguntas está efectivamente a ser respondida. Pretende-se, apesar de tudo, aproximar o parâmetro θ que nos dá a proporção de pessoas do tipo **Q** existentes na população. Supomos que a cultura geral da população inquirida permite admitir que a probabilidade de resposta sim à pergunta alternativa é igual a 0.05 (isto é, 5% da população sabe quem foi Monet)

a) Mostre que a probabilidade de obter a resposta *sim* é $\theta \times \frac{2}{3} + 0.05 \times \frac{1}{3}$.

b) Suponha que a aplicação do questionário a 100 pessoas produziu 22 *sins*. Que aproximação propõe para a probabilidade de obter resposta *sim*?

c) Que aproximação propõe para a proporção de pessoas do tipo **Q** na população, isto é, para o parâmetro θ?

109. Considere a seguinte amostra

0.892	0.249	0.410	0.017	0.908	0.008	0.389	0.032	0.001	0.001
0.210	0.678	0.036	0.006	0.374	0.001	0.012	0.253	0.084	0.001

O seu histograma sugere que é razoável admitir que a densidade da variável aleatória subjacente, X, é da forma $f(x) = (\theta + 1)\, x^{\theta} \mathbb{I}_{[0,1]}(x)$, onde $\theta > -1$. Calcule duas aproximações para a probabilidade de X ser maior do que 0.5

110. Suponhamos que temos dois estimadores d_1 e d_2 para um mesmo parâmetro θ. Se $\mathrm{E}(d_1) = \theta$, $\mathrm{Var}(d_1) = 6$, $\mathrm{E}(d_2) = \theta + 2$ e $\mathrm{Var}(d_2) = 2$, qual o estimador que devemos preferir? E se $\mathrm{Var}(d_1) = 10$?